Genome-Wide Association Analysis of Midline Axon Guidance in the *Drosophila melanogaster* Genetic Reference Panel

Undergraduate Research Thesis

Presented in partial fulfillment for the requirements for graduation "with Honors Research Distinction in Molecular Genetics" in the undergraduate colleges of the Ohio State University

by
Maya Gosztyla

The Ohio State University
April 2018

Project advisor: Dr. Mark Seeger, Department of Molecular Genetics

Introduction

The central nervous system (CNS) midline is an important choice point for many pathfinding axons during neural development. Previous studies have searched for novel regulators using mutagenesis experiments involving a few inbred laboratory strains of the fruit fly, *Drosophila melanogaster*. However, no studies thus far have attempted to utilize the polymorphic variation that exists in natural populations to study embryonic axon guidance at the CNS midline. This approach was recently enhanced by the creation of the *D. melanogaster* Genetic Reference Panel (DGRP), which consists of more than 200 isogenic, sequenced lines derived from an outbred population. In the present study, embryos from 154 DGRP lines were collected and their CNS midlines visualized using immunohistochemistry. We identified 50 lines where at least one embryo showed one or more defects in axon guidance. Next, we ran a genome-wide association analysis using two independent pipelines. We identified 6 polymorphic variants that were significant at the genome-wide threshold ($p < 1.49 \times 10^{-7}$). In addition, 27 variants were below a suggestive threshold of $p < 10^{-6}$, and of these, 10 variants were identified by both pipelines. These 27 variants are located within 16 unique genes, of which three have been previously linked to axon guidance. These results demonstrate that natural variation exists among genes influencing midline axon guidance in *D. melanogaster*. Furthermore, this work identifies novel candidates for axon guidance genes that may be investigated by functional validation in the future.

Introduction

The fruit fly *Drosophila melanogaster* is one of the most well-studied organisms in biology and serves as an animal model for countless genetic and cellular processes. Most research on *D. melanogaster* is performed on a small number of inbred strains that are shared between laboratories. This system serves to homogenize studies of diverse phenotypes and

facilitate the detection of variants. However, these inbred strains have a downside, in that they fail to account for the genetic variation that exists between natural populations of *D. melanogaster*. Traditional enhancer and suppressor screens involving these laboratory strains typically only identify relevant genes one at a time, and might miss mutations that only have observable phenotypic effects when combined together, such as due to functional redundancy or epistatic effects.

To address this problem, in 2012 a group of researchers created the *D. melanogaster* Genetic Reference Panel (DGRP), which includes 205 isogenic lines derived from an outbred natural population in Raleigh, North Carolina (W. Huang et al., 2014; Mackay et al., 2012) (Figure 1). The genome of each line is fully sequenced, facilitating the use of genome-wide association (GWA) analyses to correlate phenotypes with genetic variants. The DGRP lines capture substantial genetic diversity, containing more than 4.6 million SNP single nucleotide polymorphisms (SNPs), 100,000 polymorphic microsatellites, and 36,000 transposable element insertion sites (Mackay et al., 2012). Since the creation of the DGRP, additional resources of isogenic *Drosophila* lines have been created. One example is the *Drosophila* Population Genomes Project (DPGP), which includes more than 600 stocks from diverse geographical regions, including Zambia and Malawi (Langley et al., 2012).

The DGRP and DPGP are powerful tools for assessing gene function, as previous studies have observed that mutations often produce varying phenotypes depending on the genetic

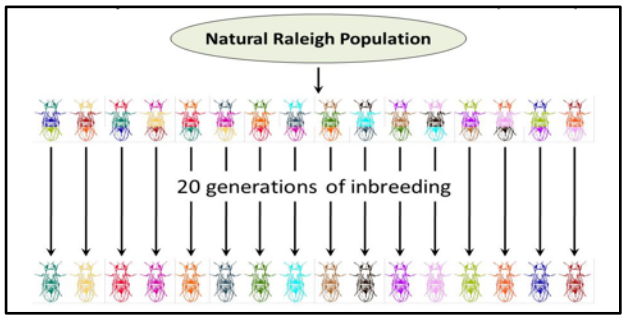


Figure 1. Generation of the *Drosophila melanogaster* Genetic Reference Panel (DGRP) from a natural population. Through separately inbreeding each line, naturally-occurring genetic variants were isolated and homozygosed. Image adapted from (Yanzhu, 2014).

background in which they are expressed (Mackay, 2014). They also allow genes previously hidden by functional redundancy or epistatic interactions to be identified, as some lines may carry unique combinations of mutations, revealing variants that individually would not have had observable phenotypic effects. Using these resources, researchers have already identified several genes with significant natural variation. For example, one such study identified extensive morphological divergence of the larval neuromuscular junction (NMJ) between DGRP lines (Campbell & Ganetzky, 2012). This insight later led to the discovery of the NMJ regulatory gene *Mob2* (Campbell & Ganetzky, 2013).

The present study applies this technique of analyzing natural variation between *D. melanogaster* strains to search for genes that are important in midline axon guidance. The human nervous system is incredibly complex, containing an estimated 100 billion neurons connected by 100 trillion synapses, each of which must be carefully guided to reach its required target. During neural development, axons are guided along specific pathways by highly-conserved guidance molecules that either attract or repel neuronal growth cones (Howard, Brown, Wadsworth, & Evans, 2017). Some of the most important guidance molecules have the job of directing growth cones at the central nervous system (CNS) midline, determining whether each axon remains on the same side of the midline or crosses over in a large bundle called a commissure (Figure 2). While the canonical axon guidance genes have largely been discovered through traditional forward genetic screens, to our knowledge there has been no attempt get to conduct a GWAS for genes involved in midline axon guidance. The DGRP and DGPG provide the necessary genetic diversity to conduct such an analysis.

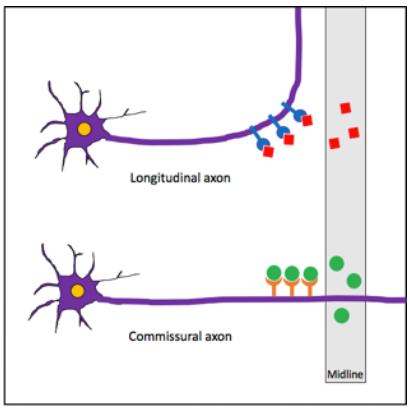


Figure 2. Axon guidance at the CNS midline. Attractive guidance molecules (green circles) direct neural growth cones to cross the midline, while repulsive guidance molecules (red square) direct the growth cones to project ipsilaterally. Both molecules are present at the midline, and the neuron's response is determined by which receptor type it expresses.

In the present study, we scored fly lines from the DGRP and the Malawi DPGP for naturally-occurring defects in midline axon guidance, and used this phenotypic data in a GWAS to identify SNPs associated with the observed defects.

Methods

Fly stocks

DGRP lines, Malawi lines, and control *white-/-* flies were obtained from the Bloomington Stock Center (Indiana, USA). Flies were maintained on corn meal media at 20°C and exposed to a 24hr light/dark cycle.

Scoring of axon guidance defects

Embryos were collected on apple juice plates and placed in multiwell mesh plates. Embryos were dechorionated in 50% bleach for 10 min, followed by fixation in 4% paraformaldehyde for 30 min at room temperature (RT). The embryos were then devitellinized using methanol and incubated in the blocking solution for 15 min. Next, embryos were incubated overnight at 4°C in one of the two primary antibodies: BP102 (mouse anti-CNS axons, DSHB, 1:20) or 1D4 (mouse anti-Fasciclin 2, DSHB, 1:10). BP102 labels all axons of the CNS, making it useful for visualizing the overall structure of the midline axon scaffold. 1D4 labels Fas2, a protein that allows growing axons to adhere to fascicles that run parallel to the midline. 1D4 is

thus an effective marker for axons that do not cross the CNS midline in wild type embryos. Following primary antibody incubation, the embryos were washed 6 times for 10 min each in 1 x PBST and then incubated in the secondary antibody (peroxidase-conjugated AffiniPure goat anti-mouse IgG, Jackson ImmunoResearch Laboratories, 1:500) for 2 hr at RT. Embryos were washed 6 times for 10 min each in 1 x PBST and then incubated for 30 min at RT in a solution of 1.5 mL 3 mg/mL 3,3'-diaminobenzidine (DAB) and 13.5 mL 1 x PBST. 150 µL of 3% hydrogen peroxide was added to the solution and incubated for an additional 45 min at RT. Embryos were washed 3 times for 10 min each in 1 x PBST and then placed in a 50% glycerol solution overnight at 4°C to clear.

Following the completion of immunostaining, embryos of the appropriate developmental stage (stage 15 for BP102, stage 17 for 1D4) were whole-mounted on a microscope slide with the ventral side facing up. Then each midline segment was visually scored under a compound microscope for the presence of defective commissures (BP102) or ectopic crossovers (1D4). Example phenotypes are shown in Figure 4A-B.

GWA analysis

Because some of the fly lines have impaired viability and/or fertility, in some cases we were not able to collect a sufficiently large quantity of embryos. Thus all lines for which fewer than 5 embryos were scored with each stain were excluded from subsequent GWA analysis. In addition, four lines that showed a severe disorganized phenotype, rather than a specific defect in axon guidance, were excluded from analysis, since these phenotypes are most likely due to a defect in early patterning of the embryo (see Figure 4, top right image). The genome data for the Malawi lines is currently only available in FASTA sequence format without variant annotation (Lack et al., 2015). For this reason, the Malawi lines were not included in the GWAS.

Phenotypes were coded as a binary trait based on the presence or absence of any defects visualized by BP102 or 1D4. Separate analyses were performed for the BP102 and 1D4 phenotypes, as well as for the presence of either phenotype. For the latter case, lines with at least one defective embryo visualized with either stain were coded as 1, lines with no defective embryos for either stain were coded as 0, and lines that did not have complete information for both stains were excluded.

Due to the unbalanced nature of our binary phenotype data (i.e., many more controls than cases), GWA analyses can be limited in statistical power. In addition, many of the most robust computational tools for GWAS rely on fitting a linear model, which requires treating the binary phenotypes as quantitative traits (for example (Vaisnav et al., 2014)). This approach is justified by recognizing the linear model as a first order Taylor approximation of the generalized linear model (Zhou, Carbonetto, & Stephens, 2013), and while it has been shown to accurately model binary traits under most circumstances, its power may be limited by unbalanced data (Cook, Mahajan, & Morris, 2017; Hayeck et al., 2015; Yang, Zaitlen, Goddard, Visscher, & Price, 2014). Therefore, to enhance the robustness of our results, we ran the analyses using two independent pipelines. The first pipeline is available on the DGRP website (http://dgrp2.gnets.ncsu.edu). This method filters out SNPs with a minor allele frequency (MAF) < 5%, and then fits a linear mixed model including a genetic relationship matrix to account to cryptic relatedness between lines. The pipeline also corrects for *Wolbachia* infection status and major chromosomal inversions (W. Huang et al., 2014).

The second independent analysis was conducted using GEMMA software (v0.97) (Zhou & Stephens, 2012). Freeze 2 genotype files were downloaded from the DGRP website. The original files contained 4,438,427 total SNPs. We first used principle component analysis to correct for the effects of population stratification. PLINK software (v1.90) (Purcell et al., 2007)

was used to clean the data by genotype missingness < 10%, MAF > 5%, and linkage disequilibrium (r^2 < 0.2, window and step size of 500 variants), as previously described by (Schmidt et al., 2017). We then used PLINK to derive the top five principle components. For the association analysis, we used GEMMA (v0.97) to apply a linear model and calculate *p*-values based on the score test. During modeling, GEMMA only included SNPs with genotype missingness < 10% and MAF > 5%. To control for cryptic relatedness and other potential confounders, we included the top five principle components, *Wolbachia* infection status, and the major chromosomal inversions as covariates. The latter two covariates were obtained from the DGRP2 website. We used the annotations from the DGRP website to annotate the significant variants using a custom R script that will be published on GitHub for free use.

The genome-wide significance threshold was selected via the Bonferroni correction based on the number of haplotype blocks in DGRP genome. This number was previously estimated by (Vaisnav et al., 2014) to be 334,729, resulting in a significance threshold of 1.49×10^{-7}. Odds ratios were calculated from the GEMMA pipeline using a log-odds transformation as described by (Pirinen , Donnelly , & Spencer, 2013). The webtool pipeline does not output an estimate for beta, so adjusted odds ratios were not computed for these data.

Results

DGRP and Malawi lines show defects in midline axon guidance

30 control embryos from an inbred *white-/-* laboratory stock were scored with each stain. None showed any defects in midline axon guidance. 154 DGRP and 6 Malawi lines were scored for axon guidance defects. A total of 2,879 individual embryos were scored, with an average of 8.6 embryos scored for each line and with each of the two stains. We identified 52 lines (50 DGRP, 2 Malawi) where at least one embryo showed one or more defects in axon guidance

(Table 1; Supp. File S1). Of these, 25 showed only BP102 defects, 20 showed only ectopic crossovers, and 7 showed both types of defects (Figure 3A). Between these 52 lines, we observed considerable variation in the penetrance of the observed phenotypes, ranging from 5% to 65% of embryos showing a defect within an individual line (Figure 3B-C).

While the appearance of the 1D4 defects were fairly consistent, the types of defects seen with the BP102 stain varied substantially (Figure 4). The most common BP102 defect seen was a thin or missing posterior commissure, which was observed for at least one embryo in 15/32 defective lines. In general, the observed phenotypes for both stains were relatively mild compared to null mutants of known axon guidance genes.

SNPs associated with midline axon guidance

To identify genes associated with axon guidance, we performed a case-control GWA analysis. Lines with at least one embryo showing an axon guidance defect were coded as cases, with the other lines as controls. The phenotypes visualized by BP102 or 1D4 were analyzed separately, in addition to a combined analysis for the presence of either phenotype. The analyses included 148, 144, and 144 lines for BP102, 1D4, and either phenotype, respectively. Malawi lines were excluded due to a lack of variant annotations.

To enhance the robustness of our results, we utilized two independent GWA pipelines to analyze the data. The first pipeline was implemented using the DGRP webtool, which included approximately 1.9 million SNPs. Our second pipeline was implemented using GEMMA and included approximately 1.6 million SNPs.

Quartile-quartile (Q-Q) plots of the GWA results indicated that the models used in the two pipelines are a strong fit for the BP102 phenotype, as well as for the presence of either phenotype (Figure 5). However, the 1D4 phenotype showed a small amount of deviation from the expected distribution for both models, suggesting that these results should be interpreted with

caution. This deviation is most likely due to the small sample size in the 1D4 GWAS, which includes fewer "cases" than the other two.

At a Bonferroni-corrected significance threshold of $p < 1.49 \times 10^{-7}$, we identified 6 SNPs that were significant for the webtool analyses and one significant SNP for the GEMMA analyses (Figure 6). Using a less conservative threshold of $p < 1 \times 10^{-6}$, we identified a total of 28 unique SNPs, of which 10 were identified by both pipelines (Tables 2-3; Figure 7A). These 10 included all 6 of the SNPs that were significant at the Bonferroni-corrected threshold. Most of the significant SNPs were located within introns (Figure 7B). The SNPs were common variants of small effect, with an average minor allele frequency (MAF) of 10.6% and average odds ratio of 1.70.

The SNPs that were not intergenic implicated 16 unique genes, including 12 annotated genes, 2 un-annotated genes, and 2 long noncoding RNAs (lncRNAs) (Table 4). Several associated genes had multiple significant SNPs within a single analysis, including ab, CG15431, Shawl, tna, and dpr12. In all cases, the multiple SNPs were in complete linkage disequilibrium with each other. The same is true for the two closely-linked intergenic SNPs identified for BP102.

Discussion

In recent years, GWAS have emerged as a powerful alternative to traditional forward genetic screens. In the present study we sought to take advantage of the genetic variation present in natural *D. melanogaster* populations to search for novel SNPs involved in midline axon guidance. To our knowledge, this is the first GWAS to be attempted in the field of axon guidance. This study adds to the growing literature suggesting that *D. melanogaster* natural population lines are a useful tool for identifying candidate genes through GWAS.

Of the 16 genes identified in this analysis, only three have been previously linked to axon guidance. Abrupt (ab) is a transcription factor that regulates genes involved in guiding motor neuron projections, though it has not previously been implicated in axon guidance at the CNS midline (Hu, Fambrough, Atashi, Goodman, & Crews, 1995). Kuzbanian (kuz) is an Adam family metalloprotease that is involved in activation of the Roundabout receptor, thereby facilitating axon repulsion from the midline (Coleman, Labrador, Chance, & Bashaw, 2010). Frazzled (fra) is a receptor for the Netrins and facilitates the ability of commissural axons to cross the CNS midline (Kolodziej et al., 1996). It is interesting that the associated phenotypes for the latter two genes seem to be counterintuitive to their known functions. While kuz null mutant embryos show ectopic midline crossovers, the kuz SNP was found to be associated with the BP102 phenotype, which is characterized by thin commissures. Similarly, fra null mutant embryos show impaired midline crossing, and yet the gene was associated with the 1D4 phenotype characterized by ectopic crossovers. A possible explanation for this may be that the associated SNPs increase the expression of kuz and fra, thus contributing to the observed phenotypes.

Among the genes with significant SNPs, four have known homologs in humans ("Homologene,"), and at least six are expressed in the embryonic CNS (Tomancak et al., 2007) (Table 4). Defective proboscis extension response 12 (dpr12), faint sausage (fas), and kekkon 5 (kek5) are notable because they contain an immunoglobulin domain, which is a common feature among many axon guidance genes (Maness & Schachner, 2007). Recent data has also suggested that the dpr family, in conjunction with dpr interacting proteins (DIPs), may regulate connectivity in a subset of neurons within in the *Drosophila* visual system (Carrillo et al., 2015; Morey, 2017; Tan et al., 2015). Myosin 31DF (Myo31DF) is also intriguing due to its role in establishing left/right symmetry within the gut (Okumura et al., 2015). The anaphase promoting

complex (APC) regulates axon growth in mammals and insects (J. Huang & Bonni, 2016). While the discoidin domain receptor (Ddr) has not been linked to axon guidance in *Drosophila*, its homologs in the nematode *C. elegans* help to guide axons along longitudinal tracts (Unsoeld, Park, & Hutter, 2013). One report in 1998 linked the faint sausage (fas) gene to axon guidance by demonstrating that fas mutant flies show a disorganized CNS, but no other publications have verified this finding (Lekven, Tepass, Keshmeshian, & Hartenstein, 1998).

Two intergenic SNPs were significant for BP102 with both pipelines. These SNPs are in complete linkage disequilibrium with each other. The webtool pipeline also identified two intergenic SNPs for 1D4. None of the closest genes to these SNPs have known roles in axon guidance or show expression in the embryonic CNS (Table 5). Therefore it appears more likely that these intergenic SNPs are exerting long-distance regulatory effects on other regions of the genome.

Future studies should confirm these results by increasing the number of identified lines with axon guidance defects. In our screen, 2 out of 6 Malawi lines from the DPGP showed a midline axon defect. This suggests that these genetic variants are not unique to North American fly populations and that larger studies will likely identify additional defective lines. In addition, since in many DGRP lines we identified only one or two defective embryos, this low penetrance may have led us to misclassify some lines as controls. Increasing our minimum number of scored embryos might identify additional lines showing axon guidance defects.

The SNPs identified in our GWAS generally had modest effect sizes. This is a common occurrence for human and fly GWAS, particularly when binary traits are analyzed (Pawitan, Seng, & Magnusson, 2009; Vaisnav et al., 2014). It is possible that the high level of functional redundancy associated with the *D. melanogaster* axon guidance pathways contributed to our low odds ratios (Howard et al., 2017). Additionally, the mild severity and highly variable penetrance

for most of the observed defects suggests that the effects of the identified SNPs can often be masked by compensatory actions of redundant genes. This is similar to the phenotypes previously observed for knockouts of axon guidance genes. For example, embryos deficient for the Netrin genes display a wide range of phenotypes and often have a substantial proportion of commissures that appear normal (Mitchell et al., 1996). This highlights the utility of GWAS for studying axon guidance, as many of these variants with subtler effects likely could not have been identified using forward genetics experiments manipulating only one gene at a time.

While insects have fewer axon guidance signaling pathways than vertebrates, the core set of pathways is highly conserved from flies to humans (Evans & Bashaw, 2010). Mutations in axon guidance genes have been linked to several human neuromuscular disorders, including horizontal gaze palsy with progressive scoliosis (Jen et al., 2004), congenital mirror movements (Srour et al., 2010), and congenital fibrosis of the extraocular muscles, Type III (Tischfield et al., 2010). Most of the relatively few axon guidance disorders identified in humans are rare. However, as pointed out by (Nugent, Kolpak, & Engle, 2012), this does not necessarily mean that problems with axon guidance do not cause human disorders, but that the technology to detect axon wiring defects in humans has only recently been created. Advances in magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI) will likely allow us to identify many additional axon guidance disorders in the years to come. Identifying genes that play a role in axon guidance may provide insight into these disorders and bring us closer to finding a cure or treatment.

Table 1. Presence (1) or absence (0) of midline axon guidance defects in the DGRP and Malawi lines. *Indicates a severe phenotype that was excluded from analysis.

Line ID	BP102	ID4
21	0	0
26		
28	1	
31	0	1
32	0	0
38	*	0
40	0	0
41		
42	1	
45		
48		
49		
57	*	0
59	0	0
69	*	
73	0	0
75	0	0
83	0	0
85	0	
88	1	
91	1	
93	0	1
100	0	1
101	0	0
105	0	0
109	0	0
129		0
136	0	1
138		
142	1	
149	1	1
153		
158		
161		
176		
177	0	0
181	0	0
189	0	0
195	0	0
208	1	0
217		
223	0	0

Line ID	BP102	ID4
227		
228	0	0
229		
233		
235	0	0
237	1	0
239	1	0
256	1	0
280	0	1
287	0	0
301	0	0
303	1	1
304	0	0
306	0	0
307	0	
309	0	0
310		
313	0	0
315	0	0
317	0	0
318	0	1
319	1	
320	0	0
321	0	0
324		0
325		
332		
335	0	0
336		
338	0	0
340		
348		
350	0	0
352	0	0
354	0	0
355		
356	0	0
357	0	0
358	1	0
359	1	0
360	0	0
361	0	0

Line ID	BP102	ID4
362	1	0
365	0	0
367	0	0
370	0	1
371		
373	0	0
374		
375	0	0
377		
379	0	0
380	0	0
381		
382	0	0
383	1	0
385	0	0
386	0	0
390		
391	0	1
392		
395		
397	0	
399	0	0
405		
406		
409		
426	0	0
427	0	1
437	0	0
439	0	0
440		
441	0	0
443	1	0
461	1	1
486	0	0
491	1	0
492		
502	1	1
505		
508	*	0
509	0	0
513	0	0
517	0	0

Line ID	BP102	ID4
528		
530	0	0
531		
535	0	0
551		
555	0	
559	0	0
563	0	0
566	0	1
584	0	0
589	0	0
595	0	0
596		
627	0	0
630	1	0
634	0	0
639		
642		
646	1	1
703	0	0
705	0	0
707	0	0
712	0	0
714	0	0
716	0	0
721	0	0
727	0	0
730		
732	0	0
737	0	0
738	0	0
748		
757	0	0
761		
765	0	1
774	0	0
776	1	1
783	0	1
786	0	0
787	0	0
790	0	0
796	0	0

Line ID	BP102	ID4
799	1	0
801	0	0
802	1	0
804		
805	0	0
808	0	1
810	0	1
812	1	0
818	1	0
819		
820	0	0
821		0
822	0	1
832	0	1
837		
843	0	0
849		
850	0	0
852	0	1
853	0	0
855	0	1
857	0	1
859	0	0
861	1	0
879	0	0
882	0	0
884		
887		
890	1	1
892	0	0
894		
897	1	0
900	0	0
907	0	0
908	0	0
911	0	0
913	0	1

Line ID	BP102	ID4
MW-6-2	0	0
MW-9-2	1	0
MW-28-2-3	0	0
MW-38-2	1	0
MW-56-2-3	0	0
MW-63-2-3	0	0

Table 2. Significant SNPs using the webtool pipeline at a significance threshold of $p < 1 \times 10^{-6}$. Italicized rows are significant at a Bonferroni-corrected threshold of $p < 1.49 \times 10^{-7}$.
Abbreviations: MAF = minor allele frequency, SS = splice site.
* Significant in both GWA pipelines.
† 2R_8433454_SNP is also located within an exon (synonymous) of the gene CG33752.

SNP	Site annotation	Variant type	Gene	MAF	*p*-value
BP102					
2L_7595046_SNP*	Intergenic	SNP		6.2%	2.34×10^{-7}
2L_7595178_SNP*	Intergenic	SNP		6.3%	2.77×10^{-7}
2L_11243596_SNP	Intron	SNP	Ab	8.3%	3.97×10^{-7}
*2L_11243721_SNP**	*Intron*	SNP	*Ab*	7.7%	*5.60×10^{-8}*
*3L_8068715_SNP**	*Intron*	SNP	*Ect4*	8.3%	*6.67×10^{-8}*
1D4					
*X_8977981_SNP**	*Intron (SS)*	SNP	*APC4*	*7.0%*	*5.72×10^{-8}*
2L_4409480_SNP	Intron	SNP	CG15431	14.1%	8.85×10^{-7}
2L_4409481_SNP	Intron	SNP	CG15431	14.2%	9.47×10^{-7}
2L_6255491_SNP	Intron	SNP	Ddr	14.0%	2.23×10^{-7}
*2L_6255512_SNP**	*Intron*	SNP	*Ddr*	*12.3%*	*1.72×10^{-8}*
2L_9409426_SNP	Intron	SNP	Shawl	12.9%	9.96×10^{-7}
2L_9409329_SNP*	Intron	SNP	Shawl	12.7%	2.76×10^{-7}
*2L_10497064_SNP**	*Intron*	SNP	*Myo31DF*	*13.5%*	*1.64×10^{-8}*
2R_8433454_SNP*†	Intron	SNP	Fra	9.9%	1.19×10^{-7}
2R_9513814_DEL	Intron	Deletion (8 bp)	Fas	8.0%	2.10×10^{-7}
2R_17485575_SNP	Intergenic	SNP		10.9%	5.83×10^{-7}
2R_19395403_SNP	Upstream (48bp)	SNP	CR43794	6.4%	6.54×10^{-7}
3L_10747400_SNP	Intergenic	SNP		5.1%	7.02×10^{-7}
3L_10857611_INS	Intron	Insertion (6 bp)	Tna	10.6%	5.02×10^{-7}
3L_10857614_SNP	Intron	SNP	Tna	10.5%	4.96×10^{-7}
3L_10857613_SNP	Intron	SNP	Tna	10.6%	5.02×10^{-7}
*3R_9303158_SNP**	*Upstream (829bp)*	SNP	*CR45589*	*7.5%*	*8.57×10^{-8}*
Either					
X_19274564_INS	Intron	Insertion (3 bp)	Kek5	8.3%	7.09×10^{-7}
2L_11243721_SNP*	Intron	SNP	Ab	7.9%	6.72×10^{-7}
2L_13633678_SNP	Intron	SNP	Kuz	20.3%	1.19×10^{-7}

Table 3. Significant SNPs using the GEMMA pipeline at a significance threshold of $p < 1 \times 10^{-6}$. Italicized rows are significant at a Bonferroni-corrected threshold of $p < 1.49 \times 10^{-7}$.
Abbreviations: MAF = minor allele frequency, SS = splice site, NS = non-synonymous.
* Significant in both GWA pipelines.
† 2R_8433454_SNP is also located within an exon (synonymous) of the gene CG33752.

GEMMA Pipeline

Variant ID	**Site annotation**	**Variant type**	**Gene**	**Odds ratio**	**MAF**	**P-value**
BP102						
2L_7595046_SNP*	Intergenic	SNP		1.43	6.2%	7.01×10^{-7}
2L_7595178_SNP*	Intergenic	SNP		1.42	6.3%	8.39×10^{-7}
2L_11243721_SNP*	Intron	SNP	Ab	1.38	7.7%	3.71×10^{-7}
3L_8068715_SNP*	Intron	SNP	Ect4	1.35	8.3%	8.56×10^{-7}
1D4						
X_8977981_SNP*	Intron (SS)	SNP	APC4	1.39	7.0%	4.23×10^{-7}
2L_6255512_SNP*	Intron	SNP	Ddr	1.31	13.5%	2.70×10^{-7}
2L_9409329_SNP*	Intron	SNP	Shawl	1.30	12.3%	4.34×10^{-7}
2L_10497064_SNP*	Intron	SNP	Myo31DF	1.30	12.7%	2.21×10^{-7}
2R_1696660_SNP	Intron	SNP	Dpr12	1.27	18.0%	5.73×10^{-7}
2R_1696677_SNP	Intron	SNP	Dpr12	1.28	19.2%	4.02×10^{-7}
2R_8433454_SNP*†	Intron	SNP	Fra	1.33	9.9%	5.24×10^{-7}
3R_7867763_SNP	Exon (NS)	SNP	CG14736	1.51	5.0%	4.96×10^{-7}
*3R_9303158_SNP**	*Upstream (829 bp)*	SNP	*CR45589*	*1.46*	*7.5%*	*1.04×10^{-7}*

Table 4. Number of significant SNPs ($p < 1 \times 10^{-6}$) for genes implicated in GWA analyses. Subsequent columns indicate whether the gene is expressed in the embryonic central nervous system (CNS) (Tomancak et al., 2007), contains an immunoglobulin (Ig) domain, or has been previously implicated in axon guidance. "?" indicates that expression data is not available.

Gene	Webtool SNPs	GEMMA SNPs	Embryonic CNS?	Ig domain?	Axon guidance?
Ab	2	1	✓		✓
APC4	1	1			
Dpr12	0	2	✓	✓	
Ddr	2	1	?		
Ect4	1	1			
Fas	1	0	?	✓	
Fra	1	1	✓		✓
Kek5	1	0	✓	✓	
Myo31DF	1	1	✓		
Shawl	2	1	?		
Tna	3	0	✓		
Kuz	1	0	✓		✓
CG14736	0	1			
CG15431	2	0	?		
CR43794	1	0	?		
CR45589	1	1	?		

Table 5. Nearest genes to significant intergenic SNPs. Distances in units of 1,000 base pairs (kbp) are based on the UCSC Genome Browser with April 2006 (BDGP R5/dm3) assembly.

SNP	Nearest upstream gene	Nearest downstream gene
2L_7595046_SNP	Spn28B (12 kbp)	Cyp4d21 (10.5 kbp)
2L_7595178_SNP	Spn28B (12 kbp)	Cyp4d21 (10.5 kbp)
2R_17485575_SNP	CG33225 (6.5 kbp)	CG10433 (8 kbp)
3L_10747400_SNP	NijA (57.5 kbp)	CG12523 (21.5 kbp)

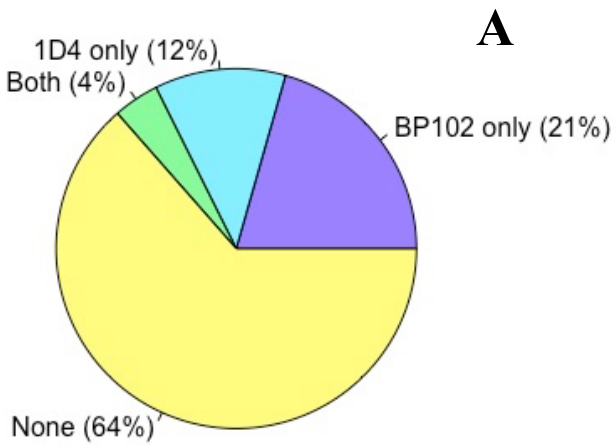
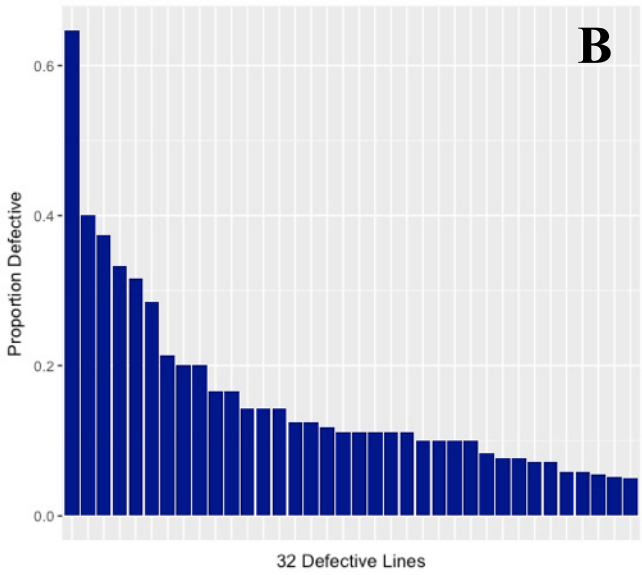
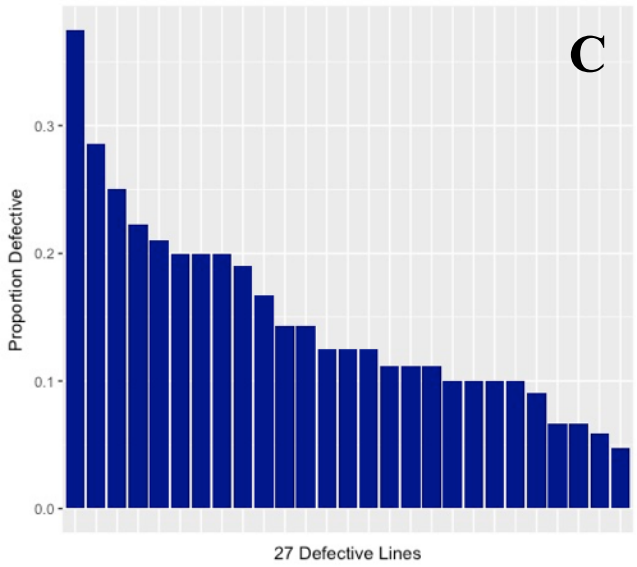
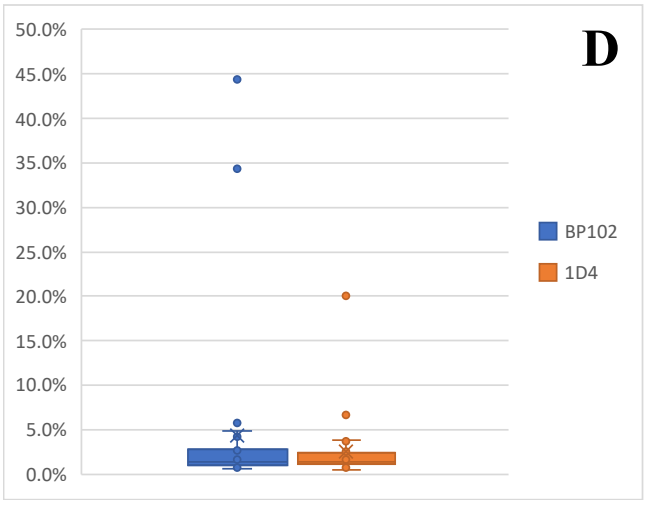
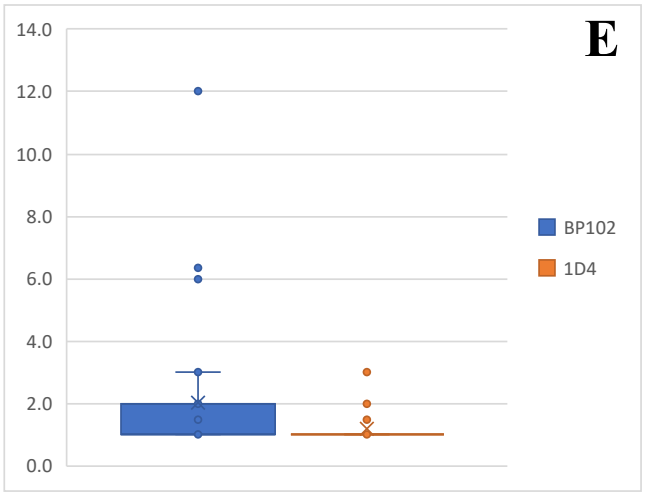


Figure 3. Midline axon guidance defects observed in the DGRP and Malawi lines. (A) 36% of the lines showed some type of axon guidance defect. Proportion of defective embryos for BP102 (B) and 1D4 (C) phenotypes. (D) Proportion of defective segments out of total segments scored in all defective lines. (E) Average number of defects per defective embryo.









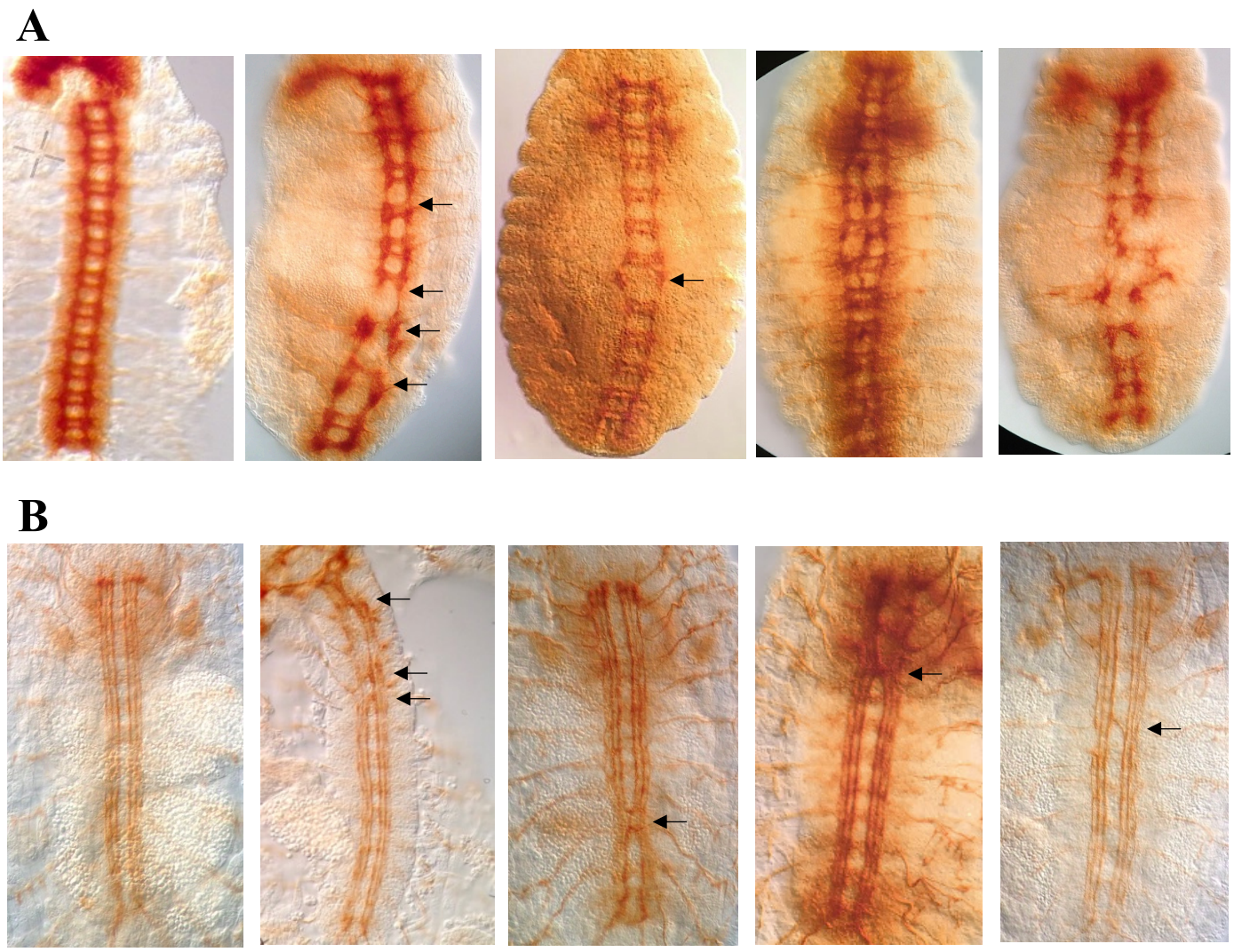


Figure 4. Types of midline axon guidance defects observed with BP102 (A) and 1D4 (B). Arrows indicate an axon guidance defect. Note that the far right image in (A) shows a severe disorganized phenotype; these lines were excluded from GWA analysis. (C) shows the types of defects observed with BP102 staining. Abbreviations: AC = anterior commissure, PC = posterior commissure, LT = longitudinal tracts.

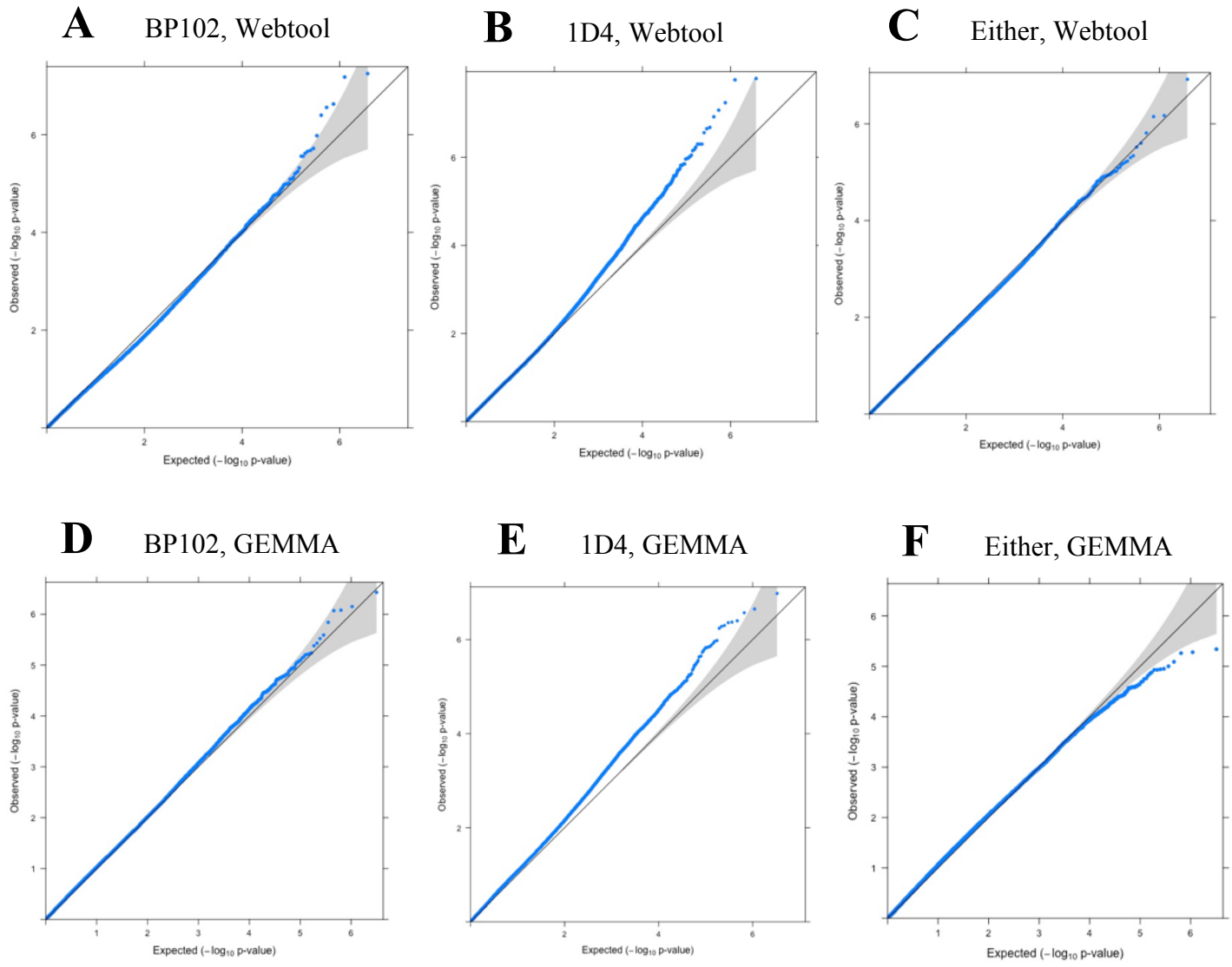
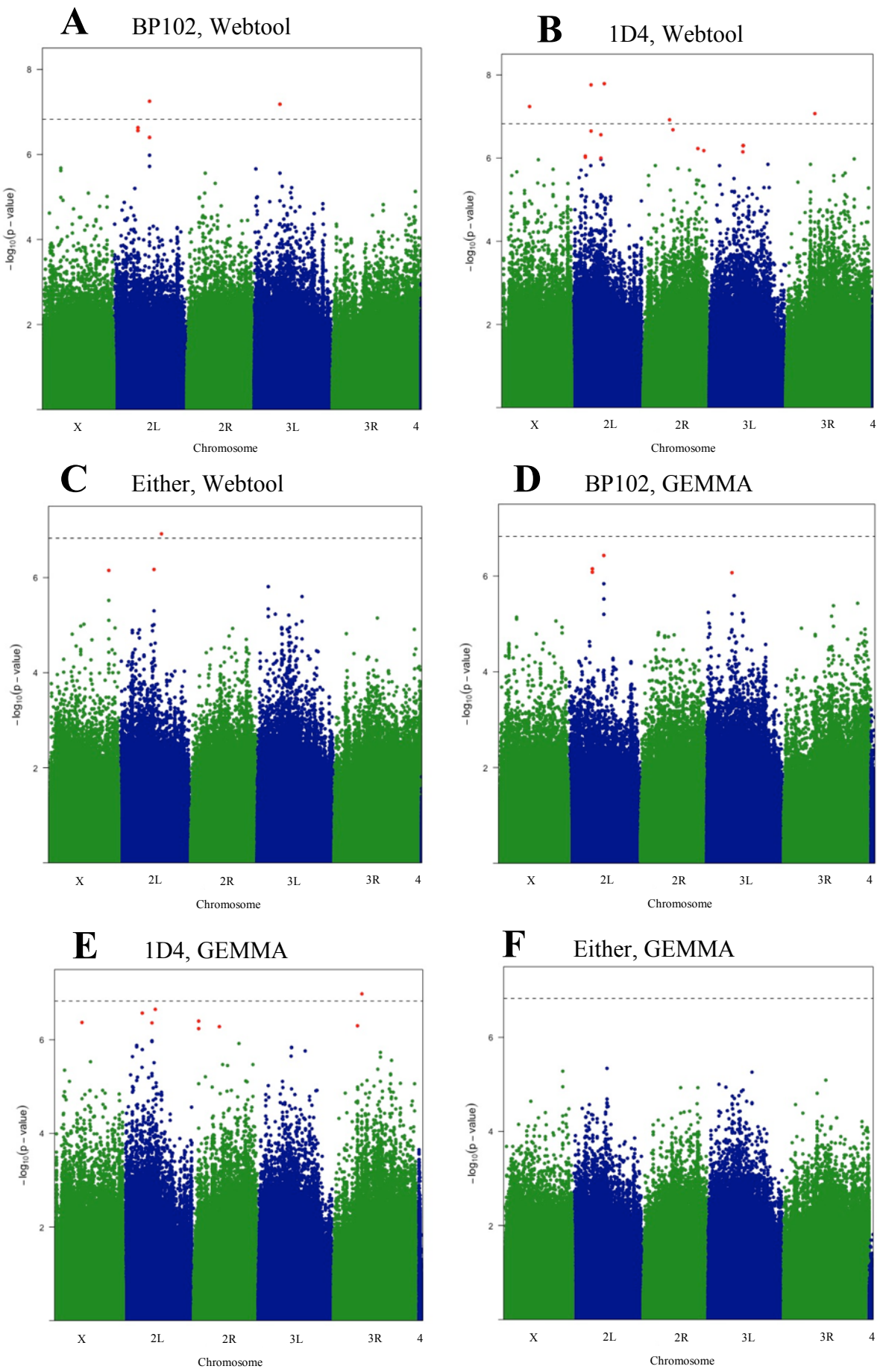


Figure 5. Quartile-quartile plots for *p*-values generated by GWA analyses for three phenotypes using two independent pipelines.

Figure 6. (below) Manhattan plots for *p*-values generated by GWA analyses for three phenotypes using two independent pipelines. The dotted line indicates a genome-wide significance threshold of $p < 1.49 \times 10^{-7}$. SNPs colored red are significance at $p < 1 \times 10^{-6}$.


A
BP102, Webtool
B
1D4, Webtool
C
Either, Webtool
D
BP102, GEMMA
E
1D4, GEMMA
F
Either, GEMMA
−log10(p − value)
Chromosome
X
2L
2R
3L
3R
4

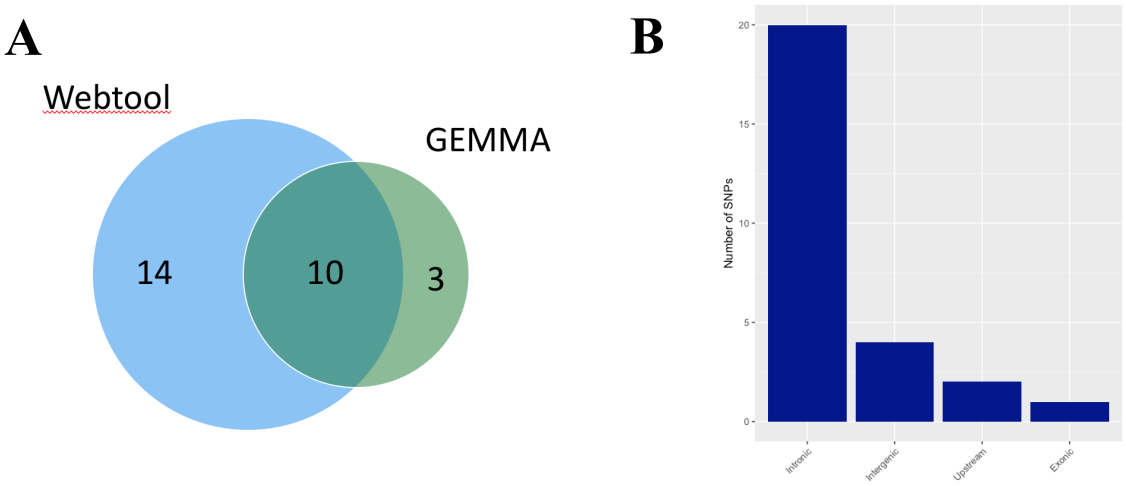


Figure 7. (A) Number of significant SNPs ($p < 1 \times 10^{-6}$) in the two GWA pipelines. (B) Locations of the 27 significant SNPs.

Acknowledgements

I wish to thank all the members of my Honors Thesis committee, particularly my mentor Dr. Mark Seeger, whose guidance allowed me to develop into a more mature scientist and whose pure enthusiasm for scientific discovery inspired me to pursue a career in research. I thank my lab-mates David Glasbrenner and Leah Anderson for the years of shared gripes, laughs, and mutual favors. Finally I thank Nick Behnke, who kept me sane through countless hours of squinting into microscopes.

This work was supported by funding from a College of Arts and Sciences Undergraduate Research Scholarship, a Mayers Summer Research Fellowship, and a Nu Rho Psi Undergraduate Research Grant.

References

Campbell, M., & Ganetzky, B. (2012). Extensive morphological divergence and rapid evolution of the larval neuromuscular junction in Drosophila. *Proc Natl Acad Sci U S A, 109*(11), E648-655. doi:10.1073/pnas.1201176109

Campbell, M., & Ganetzky, B. (2013). Identification of Mob2, a novel regulator of larval neuromuscular junction morphology, in natural populations of Drosophila melanogaster. *Genetics, 195*(3), 915-926. doi:10.1534/genetics.113.156562

Carrillo, R. A., Özkan, E., Menon, K. P., Nagarkar-Jaiswal, S., Lee, P. T., Jeon, M., . . . Zinn, K. (2015). Control of Synaptic Connectivity by a Network of Drosophila IgSF Cell Surface Proteins. *Cell, 163*(7), 1770-1782. doi:10.1016/j.cell.2015.11.022

Coleman, H. A., Labrador, J. P., Chance, R. K., & Bashaw, G. J. (2010). The Adam family metalloprotease Kuzbanian regulates the cleavage of the roundabout receptor to control axon repulsion at the midline. *Development, 137*(14), 2417-2426. doi:10.1242/dev.047993

Cook, J. P., Mahajan, A., & Morris, A. P. (2017). Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur J Hum Genet, 25*(2), 240-245. doi:10.1038/ejhg.2016.150

Evans, T. A., & Bashaw, G. J. (2010). Axon guidance at the midline: of mice and flies. *Curr Opin Neurobiol, 20*(1), 79-85. doi:10.1016/j.conb.2009.12.006

Hayeck, T. J., Zaitlen, N. A., Loh, P. R., Vilhjalmsson, B., Pollack, S., Gusev, A., . . . Price, A. L. (2015). Mixed model with correction for case-control ascertainment increases association power. *Am J Hum Genet, 96*(5), 720-730. doi:10.1016/j.ajhg.2015.03.004

Homologene. from National Library of Medicine (US), National Center for Biotechnology Information https://www.ncbi.nlm.nih.gov/homologene

Howard, L. J., Brown, H. E., Wadsworth, B. C., & Evans, T. A. (2017). Midline axon guidance in the Drosophila embryonic central nervous system. *Semin Cell Dev Biol*. doi:10.1016/j.semcdb.2017.11.029

Hu, S., Fambrough, D., Atashi, J. R., Goodman, C. S., & Crews, S. T. (1995). The Drosophila abrupt gene encodes a BTB-zinc finger regulatory protein that controls the specificity of neuromuscular connections. *Genes Dev, 9*(23), 2936-2948.

Huang, J., & Bonni, A. (2016). A decade of the anaphase-promoting complex in the nervous system. *Genes Dev, 30*(6), 622-638. doi:10.1101/gad.274324.115

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ràmia, M., Tarone, A. M., . . . Mackay, T. F. (2014). Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Res, 24*(7), 1193-1208. doi:10.1101/gr.171546.113

Jen, J. C., Chan, W. M., Bosley, T. M., Wan, J., Carr, J. R., Rüb, U., . . . Engle, E. C. (2004). Mutations in a human ROBO gene disrupt hindbrain axon pathway crossing and morphogenesis. *Science, 304*(5676), 1509-1513. doi:10.1126/science.1096437

Kolodziej, P. A., Timpe, L. C., Mitchell, K. J., Fried, S. R., Goodman, C. S., Jan, L. Y., & Jan, Y. N. (1996). frazzled encodes a Drosophila member of the DCC immunoglobulin subfamily and is required for CNS and motor axon guidance. *Cell, 87*(2), 197-204.

Lack, J. B., Cardeno, C. M., Crepeau, M. W., Taylor, W., Corbett-Detig, R. B., Stevens, K. A., . . . Pool, J. E. (2015). The Drosophila genome nexus: a population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. *Genetics, 199*(4), 1229-1241. doi:10.1534/genetics.115.174664

Langley, C. H., Stevens, K., Cardeno, C., Lee, Y. C., Schrider, D. R., Pool, J. E., . . . Begun, D. J. (2012). Genomic variation in natural populations of Drosophila melanogaster. *Genetics, 192*(2), 533-598. doi:10.1534/genetics.112.142018

Lekven, A. C., Tepass, U., Keshmeshian, M., & Hartenstein, V. (1998). faint sausage encodes a novel extracellular protein of the immunoglobulin superfamily required for cell migration and the establishment of normal axonal pathways in the Drosophila nervous system. *Development, 125*(14), 2747-2758.

Mackay, T. F. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet, 15*(1), 22-33. doi:10.1038/nrg3627

Mackay, T. F., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., . . . Gibbs, R. A. (2012). The Drosophila melanogaster Genetic Reference Panel. *Nature, 482*(7384), 173-178. doi:10.1038/nature10811

Maness, P. F., & Schachner, M. (2007). Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nat Neurosci, 10*(1), 19-26. doi:10.1038/nn1827

Mitchell, K. J., Doyle, J. L., Serafini, T., Kennedy, T. E., Tessier-Lavigne, M., Goodman, C. S., & Dickson, B. J. (1996). Genetic analysis of Netrin genes in Drosophila: Netrins guide CNS commissural axons and peripheral motor axons. *Neuron, 17*(2), 203-215.

Morey, M. (2017). Dpr-DIP matching expression in Drosophila synaptic pairs. *Fly (Austin), 11*(1), 19-26. doi:10.1080/19336934.2016.1214784

Nugent, A. A., Kolpak, A. L., & Engle, E. C. (2012). Human disorders of axon guidance. *Curr Opin Neurobiol, 22*(5), 837-843. doi:10.1016/j.conb.2012.02.006

Okumura, T., Sasamura, T., Inatomi, M., Hozumi, S., Nakamura, M., Hatori, R., . . . Matsuno, K. (2015). Class I myosins have overlapping and specialized functions in left-right asymmetric development in Drosophila. *Genetics, 199*(4), 1183-1199. doi:10.1534/genetics.115.174698

Pawitan, Y., Seng, K. C., & Magnusson, P. K. (2009). How many genetic variants remain to be discovered? *PLoS One, 4*(12), e7969. doi:10.1371/journal.pone.0007969

Pirinen , M., Donnelly , P., & Spencer, C. C. A. (2013). Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *Annals of Applied Statistics, 7*(1), 369-390. doi:10.1214/12-AOAS586

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575. doi:10.1086/519795

Schmidt, J. M., Battlay, P., Gledhill-Smith, R. S., Good, R. T., Lumb, C., Fournier-Level, A., & Robin, C. (2017). Insights into DDT Resistance from the. *Genetics, 207*(3), 1181-1193. doi:10.1534/genetics.117.300310

Srour, M., Rivière, J. B., Pham, J. M., Dubé, M. P., Girard, S., Morin, S., . . . Rouleau, G. A. (2010). Mutations in DCC cause congenital mirror movements. *Science, 328*(5978), 592. doi:10.1126/science.1186463

Tan, L., Zhang, K. X., Pecot, M. Y., Nagarkar-Jaiswal, S., Lee, P. T., Takemura, S. Y., . . . Zipursky, S. L. (2015). Ig Superfamily Ligand and Receptor Pairs Expressed in Synaptic Partners in Drosophila. *Cell, 163*(7), 1756-1769. doi:10.1016/j.cell.2015.11.021

Tischfield, M. A., Baris, H. N., Wu, C., Rudolph, G., Van Maldergem, L., He, W., . . . Engle, E. C. (2010). Human TUBB3 mutations perturb microtubule dynamics, kinesin interactions, and axon guidance. *Cell, 140*(1), 74-87. doi:10.1016/j.cell.2009.12.011

Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., . . . Rubin, G. M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biol, 8*(7), R145. doi:10.1186/gb-2007-8-7-r145

Unsoeld, T., Park, J. O., & Hutter, H. (2013). Discoidin domain receptors guide axons along longitudinal tracts in C. elegans. *Dev Biol, 374*(1), 142-152. doi:10.1016/j.ydbio.2012.11.001

Vaisnav, M., Xing, C., Ku, H. C., Hwang, D., Stojadinovic, S., Pertsemlidis, A., & Abrams, J. M. (2014). Genome-wide association analysis of radiation resistance in Drosophila melanogaster. *PLoS One, 9*(8), e104858. doi:10.1371/journal.pone.0104858

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet, 46*(2), 100-106. doi:10.1038/ng.2876

Yanzhu, L. (2014). Statistical Analysis Using RNA-Seq Data from 726 Individual Drosophila [Slideshow]. In. http://slideplayer.com/slide/5292467/.

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet, 9*(2), e1003264. doi:10.1371/journal.pgen.1003264

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet, 44*(7), 821-824. doi:10.1038/ng.2310